



Final Report
DRAFT

**DEVELOPMENT OF EASTERN REGIONAL WIND RESOURCE
AND WIND PLANT OUTPUT DATA SETS**

Subcontract No. ACO-8-88500-01

Submitted by:

AWS TRUEWIND, LLC
263 NEW KARNER ROAD
ALBANY, NEW YORK

Michael Brower
Chief Technical Officer
Tel: 978-749-9591
Fax: 978-749-9713
mbrower@awstruewind.com

November 19, 2008

Table of Contents

1	Introduction.....	1
2	Develop Wind Resource Datasets Based on Mesoscale Modeling	2
2.1	Mesoscale Model Validation	2
2.2	Mesoscale Runs	7
3	Selection of Wind Sites.....	10
3.1	Onshore Sites	10
3.2	Offshore Sites	14
4	Generate Wind Plant Output.....	15
4.1	Conversion Procedure.....	15
4.2	Validation	18
5	Forecasts	19
5.1	Forecast Synthesis Procedure	19
5.2	Validation	20
6	One-Minute Output.....	22

DEVELOPMENT OF EASTERN REGIONAL WIND RESOURCE AND WIND PLANT OUTPUT DATA SETS

Subcontract No. ACO-8-88500-01

Draft Final Report

1 Introduction

In March 2008, AWS Truewind was engaged by the National Renewable Energy Laboratory (NREL) to develop a set of wind resource and plant output data for the eastern United States. (The study region is shown in Figure 1.1.) The objective of this project was to provide wind resource inputs to the Eastern Wind Integration and Transmission Study (EWITS), an NREL-led effort to evaluate the potential impacts on the electric power system of the deployment of wind power plants meeting up to 30% of retail electric energy sales in the Joint Coordinated System Plan region (MISO/PJM/SPP/TVA) region. To carry out this study, NREL required a set of data that would capture in a realistic fashion both the temporal and spatial variability of the wind resource and associated wind power generation of onshore and offshore projects totaling 300 GW. These data were to be based on high-resolution simulations of the historical climate performed by a mesoscale numerical weather prediction (NWP) model covering 2004 to 2006.

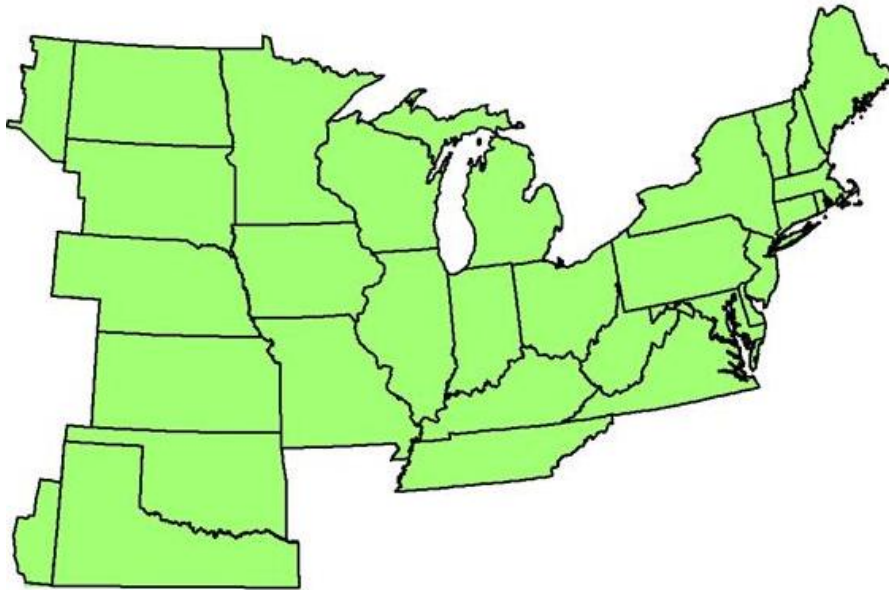


Figure 1.1. Eastern Wind Integration and Transmission Study (EWITS) region.

AWS Truewind performed this work over a period of nine months from March 2008 to November 2008. The work was divided into the following five technical tasks (as described in the Statement of Work¹):

- Develop wind resource datasets based on mesoscale modeling

¹ National Renewable Energy Laboratory, RFP RCO-8-88500, Appendix A: Statement of Work, 6 December 2007.

- Participate in the selection of wind sites
- Generate wind plant output
- Simulate forecasts
- Simulate one-minute samples of wind generation

This document presents AWS Truewind’s final technical report on the methods used and results achieved for each of these tasks. A separate validation report is provided as an appendix.

2 Develop Wind Resource Datasets Based on Mesoscale Modeling

This task was divided into two subtasks: validation of the mesoscale model and selection of the “best” model configuration and the mesoscale modeling (main runs).

2.1 Mesoscale Model Validation

The objective of the validation portion of this task was to determine the best mesoscale model and model configuration for performing the main runs. This was achieved by testing a number of different configurations covering one full year of simulations and comparing the results with wind measurements from ten tall towers in the region.

2.1.1 Validation Sites and Method

A preliminary list of seven tall towers with good data recovery in the three study years was identified by AWS Truewind. Data for the towers were extracted from AWS Truewind’s archives and time-matched. Twenty-six two-week periods in which data recovery was high at all seven towers were identified. Subsequently, at NREL’s request, three additional towers (including one offshore) were chosen to provide better geographic coverage, bringing the total number to ten. For the offshore tower we did not have data for all three years, so a single year was used. Table 2.1 lists the periods.

The ten validation towers are located in the following general areas:

- Maine (inland forested ridge top)
- New York (upstate, mixed forest/cropland, complex terrain)
- Kentucky (eastern mixed open/forested mountaintop)
- West Virginia (mixed open/forested ridgeline)
- Texas (open mesa)
- Indiana (northwest, rolling hills, cropland)
- Minnesota (south-central, open, flat)
- North Dakota (eastern escarpment, open)
- Kansas (west-central, open)
- Offshore (Atlantic)

As agreed with NREL, no further information about these masts will be provided to protect the confidential client information.

Table 2.1. Dates of simulations for the ten validation sites.

Sites	Dates		
NY, WV, IN, KY, SD, KS, ME, TX, MN	20040114-20040128	20041028-20041111	20051125-20051209
	20040128-20040211	20041223-20041231	20051209-20051223
	20040311-20040325	20050225-20050311	20051231-20060114
	20040423-20040507	20050325-20050409	20060211-20060225
	20040618-20040630	20050507-20050521	20060409-20060423
	20040811-20040825	20050604-20050618	20060521-20060604
	20041006-20041014	20050714-20050728	20050922-20051006
	20041014-20041028	20050728-20050811	20060630-20060714
	20040825-20040908	20050908-20050922	20061211-20061225
OFFSHORE ATLANTIC	20031231-20040114	20040506-20040520	20040909-20040923
	20040114-20040128	20040520-20040603	20040923-20041007
	20040128-20040211	20040603-20040617	20041007-20041021
	20040211-20040225	20040617-20040701	20041021-20041104
	20040225-20040311	20040701-20040715	20041104-20041118
	20040311-20040325	20040715-20040729	20041118-20041202
	20040325-20040408	20040729-20040812	20041202-20041216
	20040408-20040422	20040812-20040826	20041216-20041230
	20040422-20040506	20040826-20040909	

Two models were selected for testing: the Mesoscale Atmospheric Simulation System (MASS), a proprietary model developed by AWS Truewind partner MESO, Inc., and the Weather Research and Forecasting (WRF) model, a community weather model. Each model was run in a variety of configurations, which depending on the model included different turbulence parameterization schemes, numbers of vertical levels, data sources for initialization of the model runs, and data sources for assimilation.

Table 2.2 lists the configurations that were tested. Unless otherwise noted, the MASS runs employed 25 vertical levels, assimilated rawinsonde data every 12 hours, and used a turbulent kinetic energy (TKE) planetary boundary layer (PBL) scheme and one-way nesting from coarser to finer grids. The WRF runs employed 28 vertical levels and used the Yonsei University (YSU) PBL scheme and two-way interactive nesting between grids. All runs employed a horizontal grid resolution of 2 km for the innermost grid. NNGR refers to the NCEP/NCAR Global Reanalysis data set, and NARR to the North American Regional Reanalysis data set. “Sfc” refers to the assimilation of surface data. MYJ refers to the Mellor-Yamada-Janjic PBL scheme. Further information about these configurations can be found in the validation report in the Appendix.

Table 2.2 Validation experiments.

Experiment	Model	Gridded Data, Res	Other
1. MASS/NNGR	MASS 6.8	NNGR, 190 km	
2. MASS/NARR		NARR, 32 km	
3. MASS/NNGR/sfc		NNGR, 190 km	Surface data
4. MASS/NNGR/35 levels		NNGR, 190 km	35 vertical levels
5. WRF/NARR	WRF 2.2.1	NARR, 32 km	
6. WF/NNGR		NNGR, 190 km	
7. WRF/NARR/MYJ		NARR, 32 km	MYJ PBL scheme

2.1.2 Results

AWS Truewind compared the simulated and observed 80 m wind speeds for each tower and model configuration. (Hours for which the observed data were missing or invalid were excluded from the comparison.) For the simulations, the values were hourly samples extracted from the runs, i.e., the state of the model was extracted at the top of each hour. For the observations, the values were hourly averages. For the purposes of this study, this distinction has no practical impact.

To give the reader a sense of the diversity of results, Figures 2.1-2.3 present graphical comparisons of the diurnal and seasonal patterns for one of the configurations (MASS/NNGR/sfc) at three locations (Kansas, Maine, Minnesota). The red line on each chart is the observed average for that hour of the day or for that month. The green line is the raw model speed averaged for the hour or month. The blue line is the model speed scaled by a constant multiplicative factor so that the model average over all records equals the observed average. This last represents the scaling that will be done to match the expected average wind speed at a wind project site.

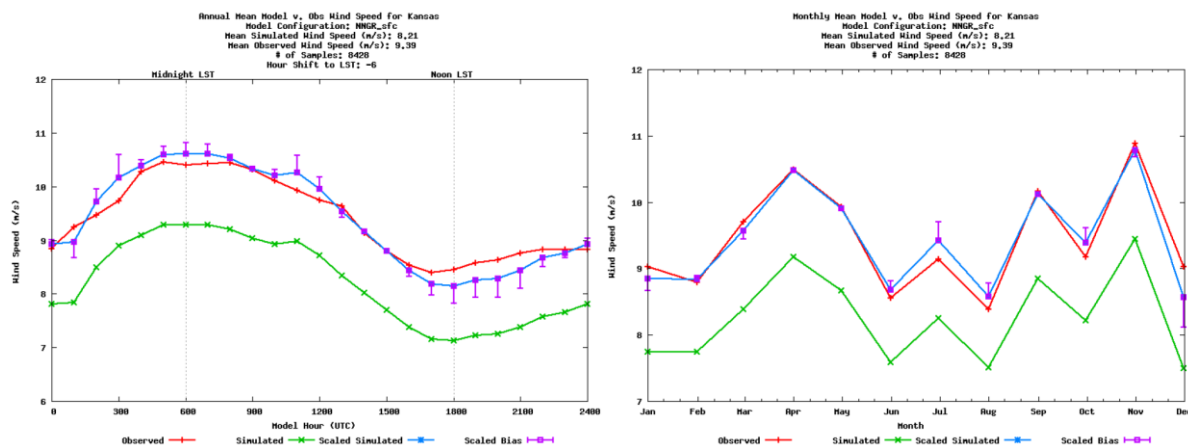


Figure 2.1. Comparison of diurnal (left) and monthly (right) patterns of average wind speed at the Kansas tower for the MASS/NNGR/sfc configuration.

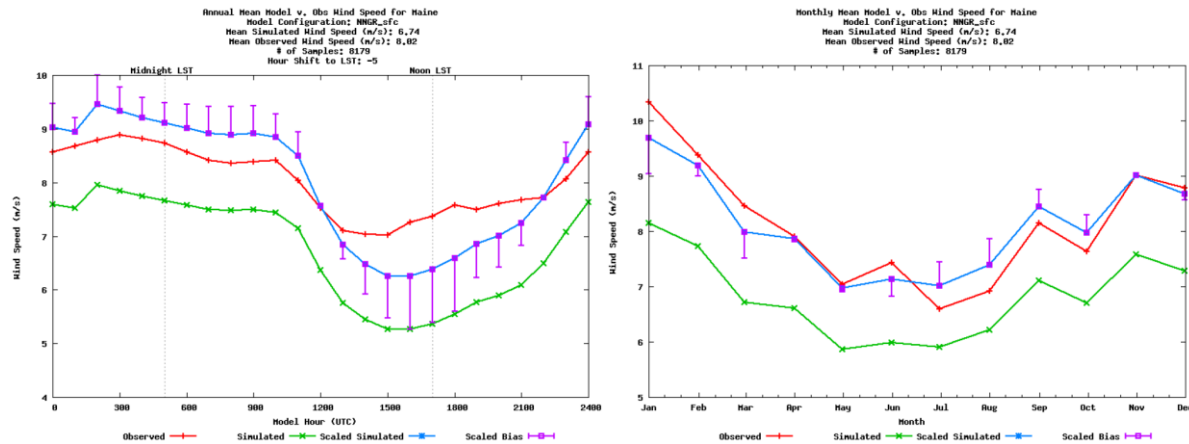


Figure 2.2. Same as Figure 2.1 for the Maine tower.

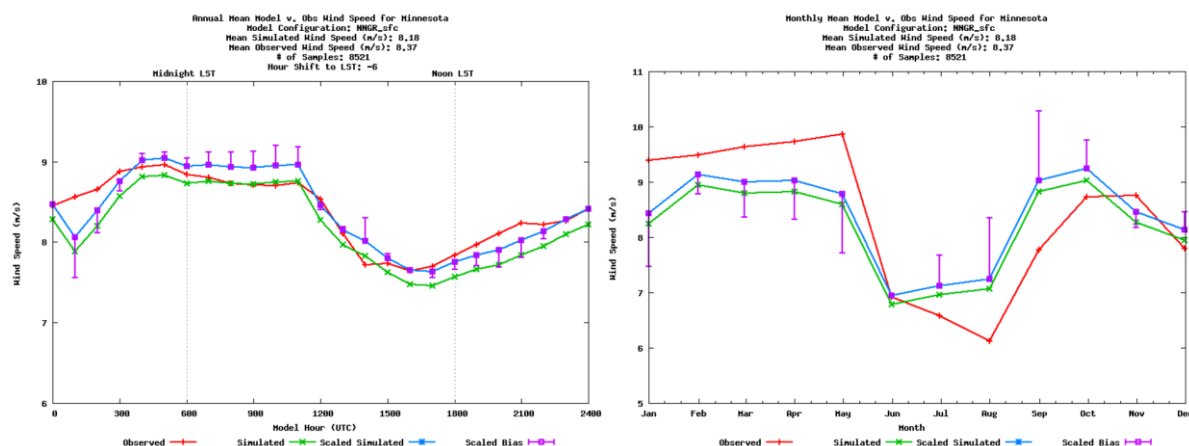


Figure 2.3. Same as Figure 2.1 for the Minnesota tower.

It was observed that at most sites all model configurations tended to underestimate the wind speed. AWS Truewind believes this is because the spatial resolution of 2 km was not adequate to fully resolve the elevated features on which most of the masts are located. Where the terrain is relatively flat (e.g., offshore), there was comparatively little bias. In addition, it was noted that at several sites the models (after scaling) tended to overestimate the wind speed at night and underestimate it during the day (especially visible in Figure 2.2). This may reflect limitations of PBL schemes, which have difficulty accurately simulating the nocturnal boundary layer. Somewhat surprisingly, the problem seems insensitive to the choice of PBL scheme. (The experiments tested three, the MASS TKE scheme and two WRF schemes.)

At NREL's request, AWS Truewind investigated certain features observed in the comparison charts: (i) It was observed that the relative model speeds of the sites in Kentucky and Maine were similar, although it would normally be expected that the Maine site would be windier. However, this was explained by the fact that the Maine site is at a much lower elevation than the Kentucky site. (ii) A marked dip in the wind speeds at some sites at 1200 UTC was ascribed to an error in the AWST data assimilation script. This was corrected for the main runs. (iii) Although the

modeled and observed monthly patterns were quite similar at most towers, there was a greater-than-normal discrepancy in some months at the Minnesota tower for all model configurations. AWS Truewind found no error in the data or analytical procedures, and also found a similar discrepancy at the nearest ASOS station. It was concluded that certain weeks in the validation data sample were poorly forecast at this location, and that the discrepancy would diminish with a larger sample of data.

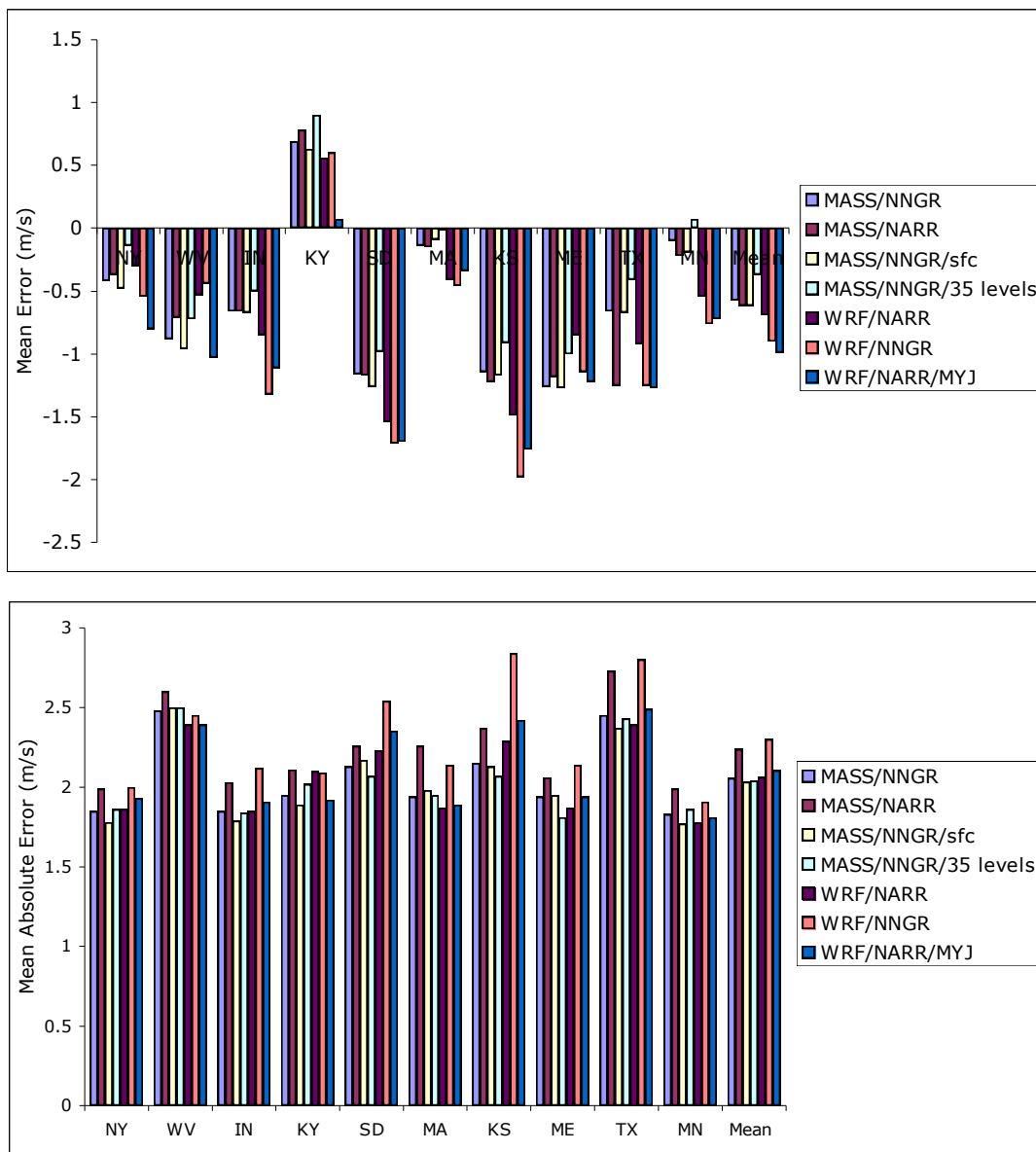


Figure 2.4. Mean error (top) and mean absolute error (bottom) over all simulations for each tower location and for all locations.

In addition to making such qualitative comparisons, AWS Truewind calculated mean error (or bias) and mean absolute error (MAE) statistics for each tower and model configuration. It was primarily on the basis of the MAE for the scaled model speeds that the best configuration was chosen. Figure 2.4 compares the annual error statistics for all the configurations.

It is evident from these charts that while some model configurations performed relatively poorly at some sites (e.g., WRF/NNGR in South Dakota, Kansas, and Texas), most configurations exhibited similar errors at most sites. This reflects the underlying similarity of the models, which solve the same physical equations. It also suggests that most errors are due not to inadequacies in the model equations (e.g., turbulence parameterization) but to an insufficient amount of observational data available for initializing and updating the simulations.

2.1.3 Findings

By a small margin, it was found that the best configuration was MASS using NNGR as the initializing data set and assimilating both surface data and rawinsonde data (i.e., MASS/NNGR/sfc).

MASS initiated with NARR data produced substantially worse results. It is suspected that this is because of an inconsistency in the definitions of land surface properties between the MASS model and the Eta model used to produce the NARR data. MASS with 35 levels did not significantly improve on MASS with the standard 25 levels, and was in some cases worse. WRF with NARR performed better than WRF with NNGR, and almost as well as MASS with NNGR. We observed little difference in the quality of the results between the MYJ and YSU PBL schemes. Unfortunately, given the tight schedule for this project, we were not able to configure WRF to assimilate rawinsonde and surface observations, as MASS is able to do. In addition, we noted that the WRF model is not able to use more than four cores on a single eight-core workstation, thus making it effectively much slower than MASS. Using WRF would have therefore entailed significant additional delays in the project schedule.

2.2 Mesoscale Runs

The main runs were carried out over a five-month period from May to September 2008. The simulations covered the period 1 January 2004 to 31 December 2006. The following table summarizes the run configuration.

Table 2.3. Model configuration for main runs

Model	MASS v. 6.8
Initialization data source	NNGR
Data to be assimilated in the course of simulations (30 km and 8 km grids only)	Rawinsonde, METAR surface observations (temperature, dew point, wind direction and speed, pressure)
Sea-surface temperatures	MODIS
High-resolution terrain and land cover (2 km grid only)	USGS 50 m CDED and 30 m LULC
Cumulus scheme (30 km and 8 km grids only)	Kain-Fritsch
Spin-up	12 hours before start of valid run
Length of run	15-16 day series (e.g., Jan 1-15, Jan 16-31)
Frequency of data sampling	Every 10 min.
Data to be stored	U, V, temperature, pressure, TKE at five heights; surface temperature and pressure, specific humidity, incoming long-wave and short-wave radiation, precipitation

The runs employed a nested grid scheme with horizontal resolutions of 30 km – 8 km – 2 km for the parent and child grids. The grid layout is shown in Figure 2.5. The largest black rectangle defines the 30 km parent grid. The three smaller black rectangles are the 8 km grids, and the small red rectangles are the 2 km grids.

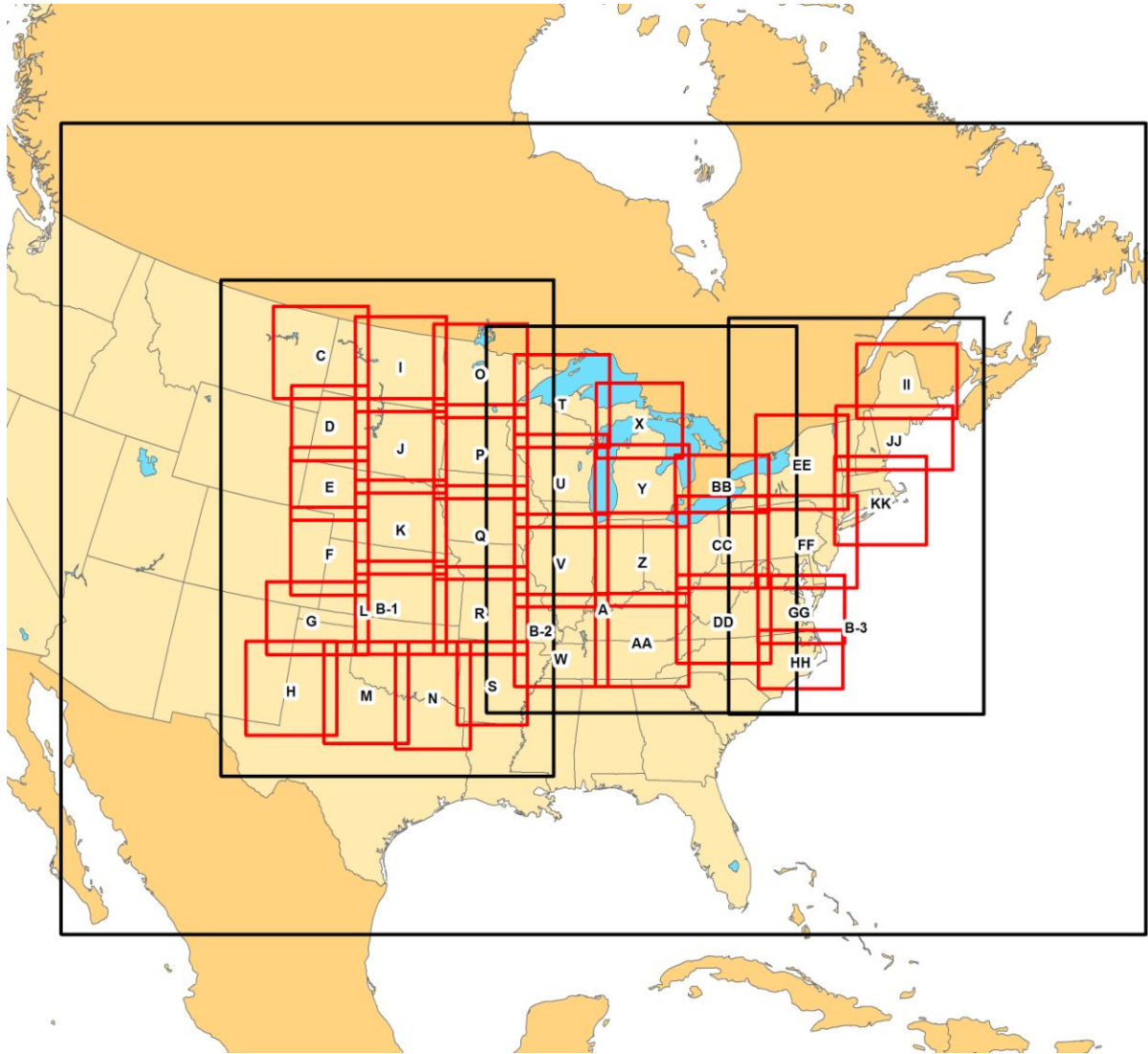


Figure 2.5. MASS grid definitions.

The simulations were carried out on a network of 640 cores on 80 dual-CPU quad-core Penguin workstations. In principle, each year should have taken about 21 days of real time on this system. However, inadequate cooling, network data congestion, and storage limits initially slowed progress to varying degrees. The cooling problem was solved with the installation of an additional 5-ton AC unit to supplement two existing 3-ton units. The network traffic issue was solved by writing more data to local scratch disks. Data storage was rectified by the installation of an additional 40 TB online RAID storage system to supplement the approximately 50 TB initially available; in addition, a high-speed tape backup system was installed to take unneeded data offline.

Once the 2004 runs were completed, three sets of mesoscale time series files were extracted for review by AWS Truewind and NREL. The three data sets consisted of: (i) grid cells near each of the 10 tall towers used in the validation and selection of the model configuration; (ii) 22 grid cells randomly selected from 11 different grids; and (iii) grid cells associated with sites near the three existing project sites for which data were provided by NREL. The data for sets (i) and (ii) were forwarded directly to NREL. The data for (iii) were used to test the power conversion program (described below).

After NREL had reviewed and approved the sample files, the data extraction was carried out for all the grid cells associated with project sites. For each grid cell, four files were produced: (i) surface pressure, (ii) 2 m temperature, (iii) 80 m speed, direction, air density, and TKE, and (iv) 100 m speed, direction, air density, and TKE. Each file for 2004 (a leap year) contains 52704 records spanning the year in 10-minute increments. The naming convention is as follows: GRID_XX_IIIJJ_HHHH.TXT, where XX refers to the grid number (from one to 36), III and JJ to the cell column and row numbers, and HHHH to the height (0000M = surface pressure, 0002M = 2 m temperature, 0080M = 80 m speed, and 0100M = 100 m speed). An example file is provided in Table 2.4. These files (which for each year number close to 120,000 and occupy about 300 GB of disk space) were not forwarded to NREL, but may be provided at a future date.

Table 2.4. Example time series file for the 80 m wind at a single grid cell. The first row indicates the latitude and longitude of the grid cell. The other rows contain the data in the following columns: (i) date in YYYYMMDD format, (ii) time (GMT) in HHMM format, (iii) speed in m/s, (iv) direction in degrees from true north, (v) air density in kg/m³, and (vi) TKE in m²/s².

47.02225	-68.80990				
20040101	0010	4.89790	270.09622	1.25625	0.03305
20040101	0020	4.94108	268.34360	1.25527	0.02336
20040101	0030	4.81025	267.33597	1.25509	0.01175
20040101	0040	4.89001	267.15210	1.25468	0.00649
20040101	0050	4.55865	265.43286	1.25415	0.00473
20040101	0100	4.70651	265.82401	1.25480	0.00252
20040101	0110	4.84289	269.14575	1.25461	0.00214
20040101	0120	4.85045	266.78668	1.25462	0.00247
20040101	0130	4.76209	266.21219	1.25440	0.00268
20040101	0140	4.74387	263.26474	1.25424	0.00220
20040101	0150	4.89790	260.24161	1.25384	0.00246
20040101	0200	4.93185	256.34119	1.25321	0.00351
20040101	0210	4.87496	252.86868	1.25324	0.00413

The 2004 runs were completed by July 2008. With space and cooling issues resolved, the pace of production accelerated, and 2005 was completed in August and 2006 in September. For each year, a set of data files similar to that shown in Table 2.4 was extracted.

3 Selection of Wind Sites

3.1 Onshore Sites

The goal of this task was to select a large number of potential wind project sites with a total rated capacity of at least 300 gigawatts (GW). Initially roughly two-thirds of the capacity was to be onshore and one-third offshore. In conversations with NREL during the early phases of the project, the desired rated capacity was increased and specific minimum targets for each state (onshore and offshore) were established.

To provide a consistent set of resource estimates for ranking and selecting sites, a seamless map of predicted mean wind speeds at 80 m height for the EWITS region was prepared from AWS Truewind's proprietary wind maps. The seamless map has a horizontal spatial resolution of 200 m, which is sufficiently fine to reflect the influence of most terrain features and to identify specific locations for wind projects. AWS Truewind has developed a method of adjusting its wind maps using a wide array of wind resource measurements to ensure good accuracy. After considerable discussion, it was decided that AWS Truewind would use a version of its maps that had been adjusted using only publicly available wind data.

A map of the estimated net capacity factor for a composite IEC Class 2 wind turbine was then created using the seamless wind speed map and speed-frequency distributions compiled from 10 years of historical mesoscale model runs previously performed by AWS Truewind at 20 km resolution. Although IEC Class 2 turbines are not suitable for every site, the use of a single curve allowed for an objective ranking of resource potential. The composite power curve was created by taking the average of three commercial megawatt-class wind turbine power curves which had been normalized to their rated capacity. The normalized average curve was then rescaled to a rated capacity of 2 MW. The composite curve is shown in Table 3.1.

Maps of exclusion areas, including parks, wetlands, and urban areas, were created as well. The site screening took into account the following exclusions:

- From USGS National Land Cover Database (2001):
 - Open Water
 - 200m buffer of Developed Low Intensity
 - 500m buffer of Developed Medium Intensity
 - 500m buffer of Developed High Intensity
 - Woody Wetlands
 - Emergent Herbaceous Wetland
- From ESRI data base:
 - Parks
 - Parks Detailed
 - Federal Lands (non – public)
 - 10,000ft buffer of small airports (all hub sizes)
 - 20,000ft buffer of large airports (hub sizes medium and large)

Table 3.1. Composite power curves for IEC Class 1, 2, and 3 turbines at standard air density (1.225 kg/m³).

Speed (m/s)	Power (kW)		
	IEC-1	IEC-2	IEC-3
0	0	0	0
1	0	0	0
2	0	0	0
3	0	0	12.6
4	39	56.6	82.4
5	136.2	176.8	204
6	280.2	347.8	378
7	474.2	574.6	621.4
8	732.6	867.8	943
9	1046.6	1213.2	1325.8
10	1404.2	1553.6	1676.6
11	1712.8	1810	1892.8
12	1911.2	1943.4	1974.2
13	1974.8	1985.2	1995.2
14	1989	1995.8	1999
15	1996.4	1999.6	1999.8
16	1998	2000	2000
17	2000	2000	2000
18	2000	2000	2000
19	2000	2000	2000
20	2000	2000	2000
21	2000	2000	2000
22	2000	2000	0
23	2000	2000	0
24	2000	2000	0
25	2000	2000	0

- Other:
 - Slopes greater than 20%
 - Areas outside the study region

At NREL's request, data from the Conservation Biology Institute were added. The excluded layers had GAPCAT values of 1, 2, 7 and 8, which are defined as areas managed to maintain a natural state or for natural values and managed for conservation to varying degrees. Many of the areas defined in the CBI data base coincided with exclusions already defined in the other data bases. In those cases, no changes in the exclusions occurred.

In the first round of site screening, over 2000 GW of sites with a maximum capacity of 2000 MW each were selected. Upon review by NREL, it was concluded that many of the sites,

particularly in the western part of the study region, were too large. At NREL's request, a second round of site screening was done, with the maximum project size set to 1000 MW. This was also deemed unsatisfactory. AWST then modified its site-screening program so that it would select sites with a range of rated capacities, even in areas where very large sites could be supported. This final approach, which produced over 7800 sites totaling over 3000 GW of rated capacity, was deemed acceptable by NREL.

Table 3.2. State totals of onshore and offshore sites. Sizes of onshore sites range from 100 MW to over 1000 MW. Each offshore site is 20 MW.

State	Onshore		Offshore	
	Count	Total MW	Count	Total MW
Arkansas	20	4049		
Colorado	8	3760		
Connecticut	7	919	84	1680
Delaware	7	1018	221	4420
Illinois	79	42029	99	1980
Indiana	63	32591	112	2240
Iowa	92	52575		
Kansas	66	46069		
Kentucky	6	1490		
Maine	42	5863	64	1280
Maryland	9	1114	496	9920
Massachusetts	19	2166	1006	20120
Michigan	57	23944	788	15760
Minnesota	121	61480		
Missouri	19	10138		
Montana	12	5830		
Nebraska	89	48471		
New Hampshire	21	2371	1	20
New Jersey	8	1328	1343	26860
New Mexico	24	10524		
New York	66	14859	502	10040
North Carolina	10	1999	2244	44880
North Dakota	60	32138		
Ohio	34	17445	1479	29580
Oklahoma	82	40253		
Pennsylvania	56	6988	257	5140
Rhode Island	7	1039	65	1300
South Dakota	91	48547		
Tennessee	8	886		
Texas	48	31896		
Vermont	17	2019		
Virginia	16	2097	1507	30140
West Virginia	18	2376		
Wisconsin	44	20494	162	3240
Total	1326	580762.5	10430	208600

An additional concern raised by NREL was the small number of sites chosen in four states: Connecticut, Rhode Island, New Jersey, and Delaware. To address this concern, AWST ran a separate site screening with a lower capacity factor threshold for those states alone, and the sites chosen were added to the others. In the last step, NREL manually selected a subset of the sites to arrive at the final site selection.

The final site list contains 1326 sites totaling 580 GW. Figure 3.1 shows the locations of the sites, color-coded by predicted annual mean capacity factor, and Table 3.2 summarizes their distribution by state. In the Great Plains, the predicted capacity factor typically exceeds 40%, whereas in most eastern states it is below 30%. Figure 3.2 shows the distribution by rated capacity. The bulk of the sites fall between 100 MW and 600 MW in size. A small number of “megasites” with rated capacities exceeding 1000 MW were also chosen. All of these are in the Great Plains.

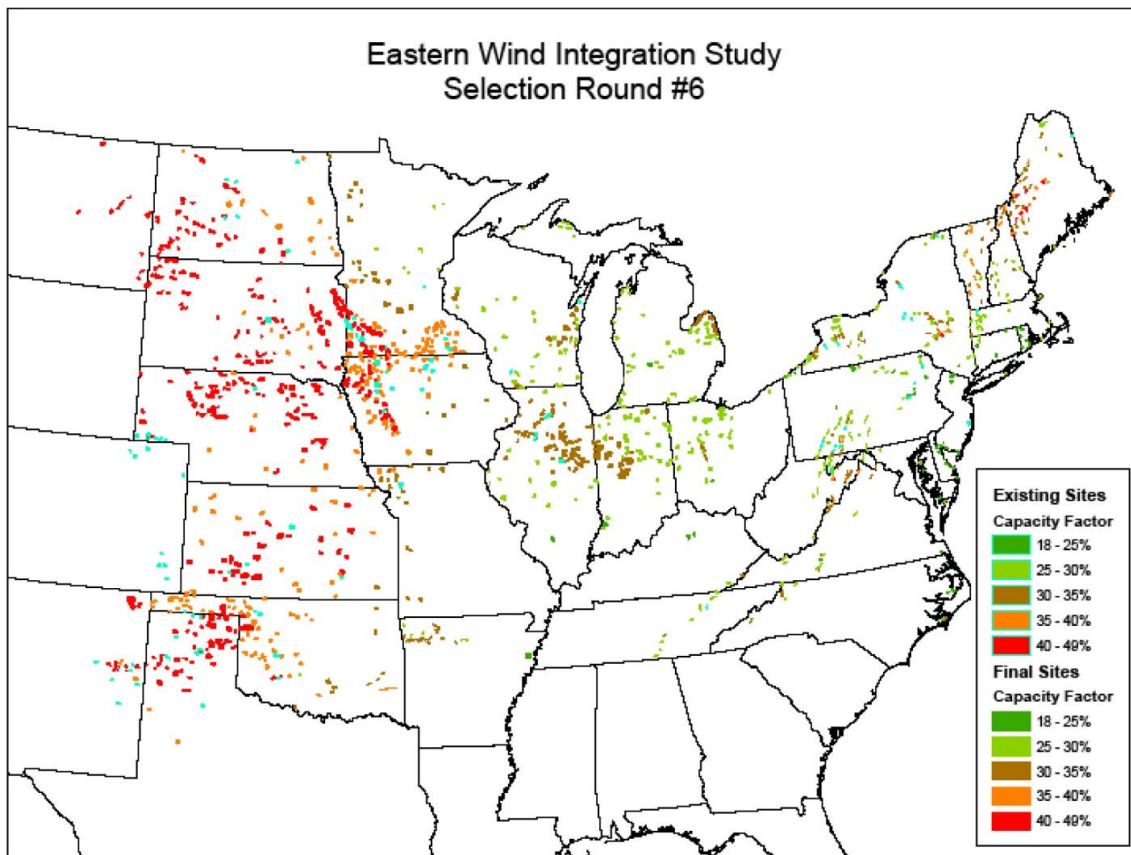


Figure 3.1. Locations of 1326 onshore sites. Light blue points show the locations of existing projects.

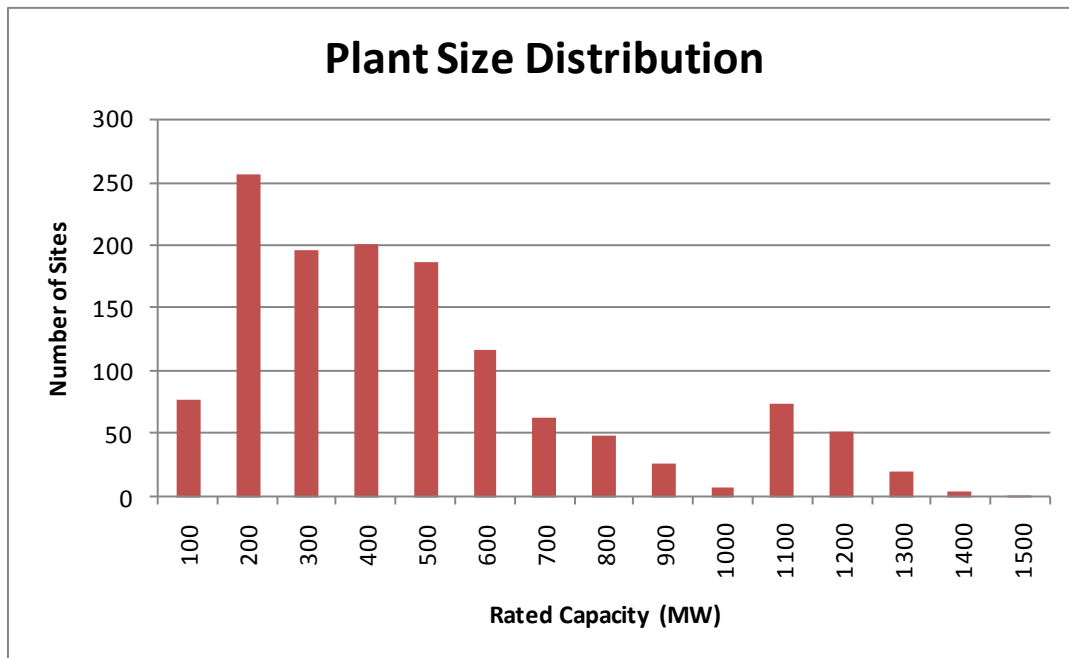


Figure 3.2. Distribution of plant sizes for 1326 onshore sites.

3.2 Offshore Sites

For the offshore site screening, NREL requested sites with a maximum water depth of 30 m and minimum distance from shore of 5 miles (8 km). In the initial round of offshore site screening, AWS Truewind selected all offshore areas meeting these criteria that were also outside federal or state protected areas and had an estimated net annual capacity factor (CF) of at least 32%. The CF was determined from the mean speed at 80 m from AWS Truewind’s wind map, speed distributions from AWS Truewind’s historical 20 km resolution simulations, and the default composite IEC Class II wind turbine power curve. (The net CF is not meant to be realistic for offshore projects but was used for ranking purposes only.) The selected areas were forwarded to NREL in shape file format.

After NREL’s review, the selected regions were converted to a list of offshore “sites.” The sites were presented somewhat differently from the onshore sites. Because there is so little spatial variation in the wind resource well offshore, rather than attempt to identify contiguous areas meeting a minimum capacity threshold such as 100 MW, it was decided to select points spaced approximately 2 km apart. Each point represents roughly a single mesoscale grid cell 4 km² in area, which is assumed to be capable of supporting 20 MW of offshore wind capacity (i.e., a mean density of 5 MW/km²).

The list included just over 16000 such points. In addition to latitude and longitude and nearest state, each point was characterized by the predicted mean speed at 80 m and 100 m and the predicted net CF at 80 m height. Included in the spreadsheet was a running sum of the gigawatts of wind capacity associated with each state, in order of decreasing net CF, and a table in a separate worksheet showing the grand total for each state.

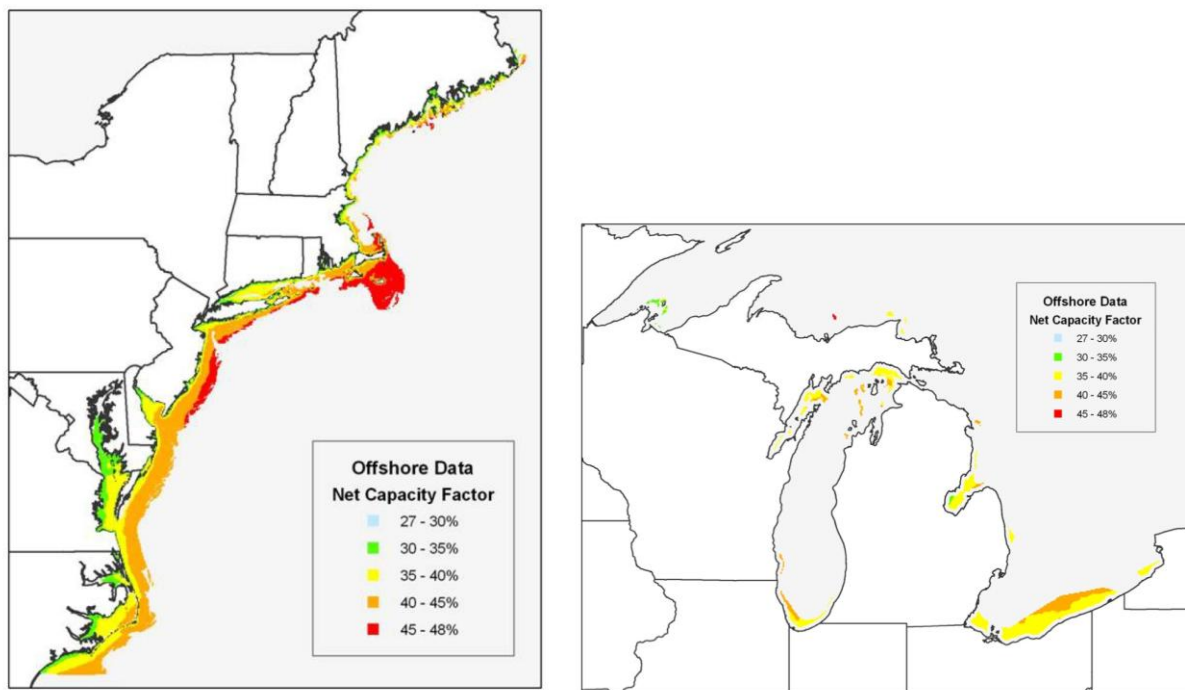


Figure 3.2. Offshore sites.

In its review, NREL detected inconsistencies between the estimated site mean speed and net CF. Some sites had a lower speed than was apparently justified by the CF. AWS Truewind found that the wrong speed map had been used when extracting the speeds for the table. In addition, AWS Truewind discovered that sites closer than 5 miles (8 km) from the Atlantic coast had been mistakenly retained in the list. Both problems were corrected, and a new table containing about 10,000 offshore sites was sent to NREL. The distribution by nearest state is shown in Table 3.2.

4 Generate Wind Plant Output

4.1 Conversion Procedure

Converting the meteorological data generated by the mesoscale model to wind plant output was done by a program written by AWS Truewind called SynOutput.

The program starts by reading a list of the 10 validation towers and their nearest associated grid cells (grid number and column and row position). It also reads a list of the grid cells associated with the sites in the 600 GW scenario. In general, several grid cells are associated with each site. For each cell, the list provides the latitude and longitude, expected mean speed, mesoscale elevation, actual elevation, and proportion of the site's total rated capacity associated with that cell. The mean speeds are based on AWS Truewind's wind maps adjusted to the year of the simulation. Since each year was produced separately, it was not possible to scale the mean speeds for all three years to the expected 2004-2006 mean at each site. AWS Truewind proposed that its 20 km historical data set be used to establish each year's mean relative to the 1997-2007

period. It is expected that this procedure will result in only minimal error. NREL accepted this proposal.

The program then imports the turbine power curves. There is one power curve for each IEC class. (The curves are shown in Table 2.1.) The power curves are scaled to a rated capacity of 2 MW and are valid for the standard sea-level air density of 1.225 kg/m^3 . The IEC 1 and 2 curves are based on a composite of three commercial turbines (GE, Vestas, Gamesa brands). In consultation with NREL, it was decided to base the IEC 3 curve on just two turbines (GE 1.5xle and Gamesa G90) to avoid an inconsistency in the cut-out speed of the Vestas V100. In addition, the cut-out speed of the GE turbine was changed from 20 to 21 m/s to match that of the Gamesa turbine. The IEC 1 and 2 turbines are assumed to have a hub height of 80 m and the IEC 3 turbine 100 m.

The program next reads a set of 12x24 speed matrixes, one for each of the 10 validation towers. These matrixes give the mean speed for each hour of the day and for each month of the year. For each tower there are two matrices, one for each hub height (80 m and 100 m). The program reads the mesoscale time series file for each of the grid cells nearest the validation towers. From the speed data it creates a 12x24 mean speed matrix for each hub height. The ratio between the average observed speed and the average simulated speed is then calculated for each bin and normalized to an average of one. The result is an adjustment matrix which is used to correct model biases.

The mesoscale time series file for each grid cell associated with a project site are then read. The speed data are scaled to match the expected mean speed and finally summed for all the grid cells associated with the site. In the sum, each cell's speeds are weighted according to the proportion of the site area associated with that cell. The result is a time series of simulated wind speeds for the site as a whole at both 80 m and 100 m.

The program calculates a correlation coefficient (r^2) between the simulated daily mean speeds for the site in question and the simulated daily mean speeds for each validation location. It then calculates a weighted average adjustment matrix for the site in which the weight given to the adjustment matrix for each validation location is proportional to its correlation coefficient. The program applies this blended adjustment matrix to the simulated data for the site. For example, if the time in question is 1300, the simulated speed is multiplied by the adjustment factor for 1300. Although the program calculates adjustments on a monthly basis, because it was found during the validation phase that the monthly variation in speeds was accurately predicted by the model, the monthly diurnal adjustment factors are ignored and only an annual adjustment is performed.

The speeds for each grid cell are then adjusted for wake losses in a manner that depends on the simulated wind direction relative to the prevailing (most frequent) direction. The loss is given by $w = w_{\min} + (w_{\max} - w_{\min}) \sin^2(\theta - \theta_{\max})$, where w_{\min} is the minimum loss (assumed to be 4%) when the wind is aligned with or opposite to the prevailing direction θ_{\max} , and w_{\max} is the maximum loss (9%) when the wind is perpendicular to the prevailing direction. The loss factor accounts both for wake losses and implicitly for other losses such as blade soiling that affect the efficiency of power conversion for a given free-stream speed without reducing the maximum output. The speed is further adjusted by adding a random factor scaled to the predicted TKE. This adjustment is intended to reflect the impact of gusts on the speeds experienced by the turbines in the wind project.

Next, the adjusted speeds are applied to the turbine power curve for each IEC class. In the process, the power curve is corrected to the predicted air density. A time filter is then applied to mimic the effect of spatial averaging on the fluctuations of wind output over the area of a mesoscale grid cell. The time filter gives a weight of 90% to the predicted output at the current time, and divides the remaining 10% weight equally among the predicted output values of the previous 17 time records (i.e., about 3 hours of actual time). This approach has been found to reproduce the observed “variability” of wind plant output, as measured by the mean absolute deviation (MAD) as a function of time, with reasonable accuracy. Without such time filtering, the simulated plant output would tend to be more variable than the observed.

The program applies an additional power loss to account for turbine and plant availability. Based on data obtained by AWS Truewind for operating wind projects, the availability is assumed to follow a normal distribution with a mean of 94.8% and a standard deviation of 2.3%; the distribution is truncated at 100%. To avoid unrealistic rapid fluctuations in output, the availability is allowed to change on average only once per hour. An additional loss of 3% is subtracted from the output to represent electrical losses. For most sites, the total loss is about 15%-17%. The range is from about 12% to about 20%.

For each site, the simulated speeds at both hub heights and the power output for all IEC classes are output to a single text file. In addition, the program selects the most appropriate IEC class based on the maximum mean speed within the site adjusted for air density. The power output for the selected IEC class is provided in the last column of the file. In this way, users of the data can simply import the last column rather than have to select themselves which IEC class to use. A sample text file is shown in Table 4.1.

Table 4.1. Sample plant output data file.

SITE NUMBER: 4403 RATED CAP: 100.2 IEC CLASS: 3 LOSSES (%): 19.5 18.6 17.0							
Date (GMT)	Time (GMT)	Speed (80m)	Speed (100m)	IEC1 (80m)	IEC2 (80m)	IEC3 (100m)	Selected
20060101	10	3.876	4.062	1.2	1.8	3.3	3.3
20060101	20	3.982	4.314	1.4	2.0	4.2	4.2
20060101	30	4.844	5.204	4.0	5.3	8.3	8.3
20060101	40	5.549	5.929	7.2	9.2	12.8	12.8
20060101	50	5.999	6.471	9.8	12.2	17.0	17.0
20060101	100	6.242	6.785	11.2	14.0	20.0	20.0
20060101	110	6.085	6.687	10.1	12.7	18.7	18.7
20060101	120	5.653	6.315	7.9	10.0	15.9	15.9
20060101	130	4.852	5.563	4.1	5.4	10.5	10.5
20060101	140	3.870	4.592	1.6	2.2	5.9	5.9
...

4.2 Validation

NREL provided three years of 10-minute plant output data for three projects within the EWITS area: Lake Benton (MN), Storm Lake (IA), and Blue Canyon (OK). (A fourth project, Trent Mesa, TX, is outside the EWITS area and was therefore not used.) The power conversion program was tested using mesoscale data extracted for grid cells associated with the three projects, and the results were provided to NREL both in tabular form and as diurnal and change plots. By using an appropriate turbine power curve for each project and by adjusting the hub height of the simulated speeds and diurnal corrections to match the actual hub height, a reasonably close agreement between the predicted and observed net capacity factors and diurnal patterns was obtained (Figure 4.1). The mean absolute changes in output as a function of time lapse, an indication of the variability of output over a range of time scales, were also compared. The curves for the simulated and observed wind plant output follow similar trajectories (Figure 4.2).

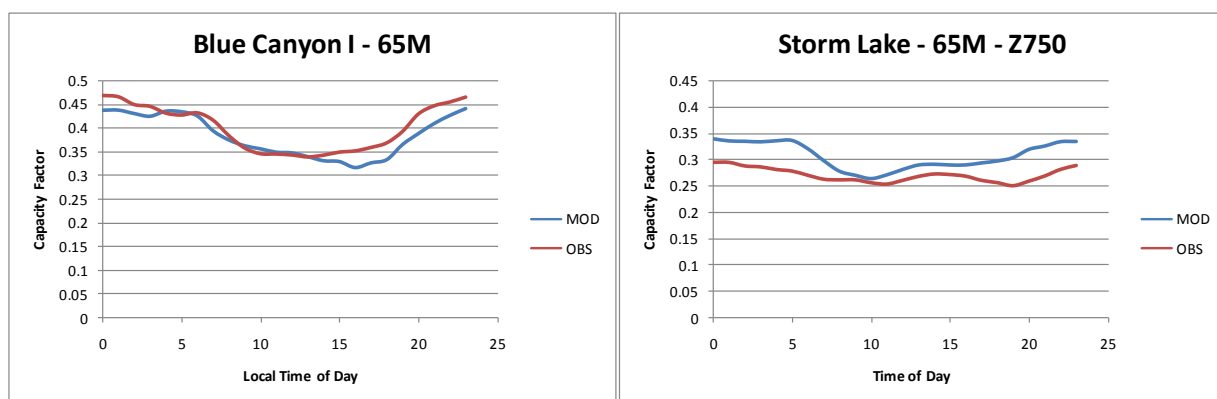


Figure 4.1. Comparison of simulated and observed diurnal mean wind speeds for two validation projects, Blue Canyon I (left) and Storm Lake (right).

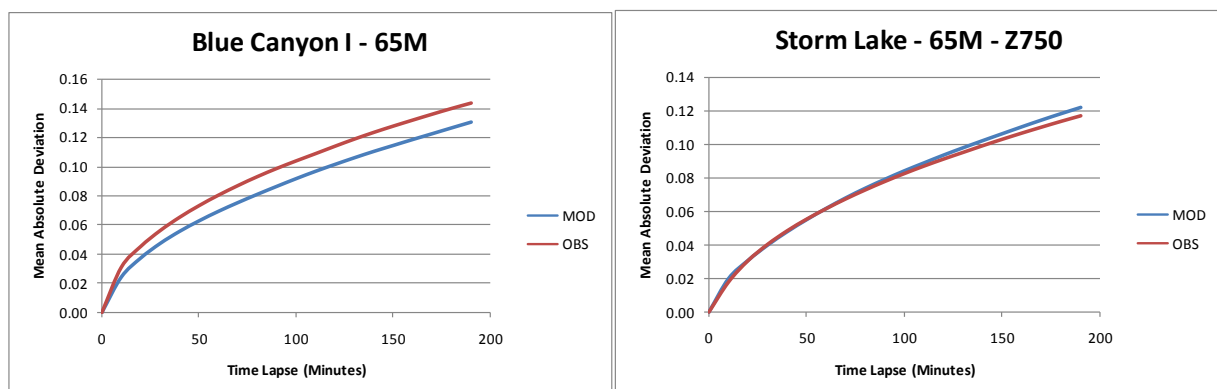


Figure 4.2. Comparison of simulated and observed mean changes in plant output as a function of time lapse for Blue Canyon I (left) and Storm Lake (right).

In reviewing the plant output files, MISO noted significant jumps at 0z and 12z on many of the days AWS Truewind concluded that the jumps were caused by the abrupt assimilation of rawinsonde and surface observations every 12 hours in the mesoscale runs. A workaround was developed, in which values spanning the affected times were replaced with synthesized data (Figure 4.3). Two other errors in the data processing were discovered at the same time, and a

new set of data was produced. With these revisions, the final plant output data files were regenerated and delivered to NREL.

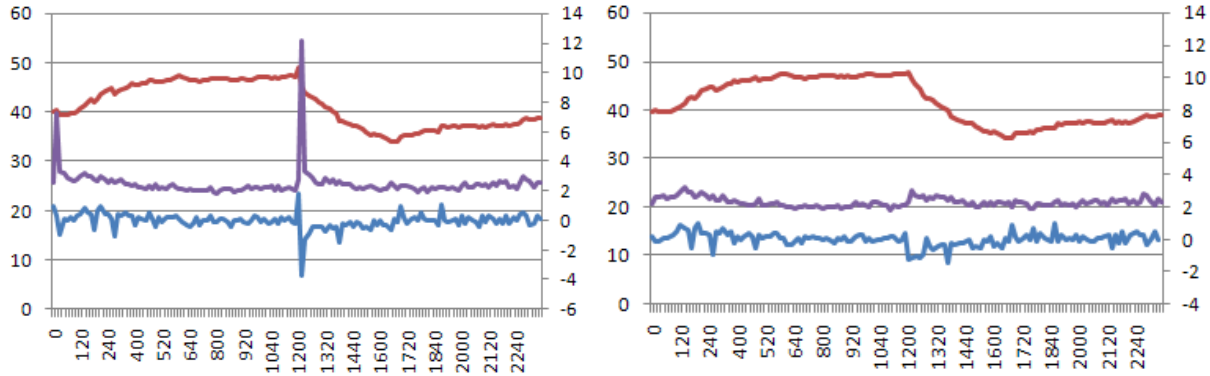


Figure 4.3. Jumps in mean speed at one site before (left) and after (right) the fix. The red curve is the mean output (left axis), the purple curve is the absolute change in output from one 10-minute record to the next (left axis), and the blue curve is the change in output (right axis).

5 Forecasts

AWS Truewind produced hourly forecasts for three different time horizons: next-day, six-hour, and four-hour. Each set of forecasts was synthesized by running a statistical forecast synthesis tool written by AWS Truewind called SynForecast. This tool uses actual forecasts and observed plant output to develop a set of transition probabilities, which are then applied stepping forward in time from a random starting point in a process known as a Markov chain. The procedure is described in depth in the following section.

5.1 Forecast Synthesis Procedure

AWS Truewind ran its eWind forecasting system in “hindcast” mode for the four wind plants for which NREL had previously provided output data for 2004-2006: Trent Mesa, Blue Canyon, Lake Benton, and Storm Lake. The mesoscale model feed was provided by 8 km resolution MASS simulations. The actual plant data up to the time the forecast was assumed to be generated provided the observed data feed. For the next-day forecasts, this was 5 pm local time of the day before the forecast; this effectively assumes the forecasts are generated once per day. For the four-hour and six-hour forecasts, the latest time was four hours and six hours ahead of the delivery time, which implies an hourly update schedule.

Since no on-site wind measurements were available, the eWind statistical module transformed the mesoscale model data directly into plant output via an implicit non-linear power curve derived from data for the preceding 30 days of actual and simulated data. The forecasts were issued in 24-hour periods from midnight to midnight of each forecast day and appended to one another to form a continuous time series of forecasts.

From each of these four sets of forecasts, the SynForecast program constructed a matrix of forecast probabilities of the following form:

$$P(A_t \cap F_{t-1} \cap F_t)$$

The probability P is the number of occurrences for which the actual output was A_i and the forecasted outputs were F_{t-1} and F_t , where t is a particular moment in time and $t-1$ is the previous moment (one hour earlier). Before constructing this matrix both the actual and forecasted output values are normalized to the rated capacity of the wind project and placed in 10 bins ranging in capacity factor from 0.05 to 0.95 in increments of 0.10. Both the current and previous forecasts are included in the probability matrix to capture the autocorrelation of forecast errors, as otherwise the synthesized forecasts would fluctuate randomly about the actual output in an unrealistic fashion.

For each wind project site, the SynForecast program selected, at random, one of the four probability matrixes. Starting with a random seed, the program stepped forward in time taking random draws from the transition matrix. In this manner, an hourly next-day forecast was synthesized exhibiting the same error distribution and autocorrelation behavior as the actual forecasts.

5.2 Validation

To verify that the program was working properly, we compared synthesized forecasts with the actual forecasts for the four validation wind projects. First we considered the time correlation of the actual and forecasted generation and the root-mean-square (RMS) forecast error. The results of the actual and synthesized forecasts were very similar, as shown in Table 5.1. The RMS error depends in part on the average plant output, with more productive plants experiencing higher forecast errors as a fraction of rated capacity because they spend more time in the steeply sloping parts of their power curves.

Next, we considered the autocorrelation of the output, the forecasts, and the forecast errors. The autocorrelation indicates the degree to which a particular parameter tends to persist over time. A parameter that typically changes little would have an autocorrelation factor of nearly one, whereas one that fluctuates randomly would exhibit an autocorrelation factor of nearly zero. The following tables indicate that the observed output of the four wind projects tends to be quite strongly autocorrelated over a period of one to several hours. The eWind and synthesized forecasts exhibit similar degrees of autocorrelation in each case. As for the forecast errors, their autocorrelation is considerably lower, and the SynForecast program seems to capture the pattern of decreasing correlation with increasing time shift quite well. However, the synthesized forecasts are slightly less correlated in time than the eWind forecasts, while the synthesized forecast errors are slightly more correlated.

Table 5.1. Comparison of correlation of forecasted and actual output and RMS forecast error for synthesized and real (eWind) forecasts.

Plant	Correlation (Pearson r)		RMS Forecast Error (CF)	
	eWind	SynFcst	eWind	SynFcst
Trent Mesa	0.77	0.77	0.20	0.20
Blue Canyon	0.77	0.73	0.21	0.22
Storm Lake	0.79	0.81	0.16	0.16
Lake Benton	0.72	0.71	0.19	0.19

Table 5.2a. Autocorrelation of actual output, eWind forecasts, and synthesized forecasts and forecast errors: Trent Mesa

Time Shift	Obs	eWind	SynFcst	eWind Error	SynFcst Error
0	1	1	1	1	1
1	0.93	0.94	0.92	0.78	0.81
2	0.85	0.87	0.85	0.59	0.65
3	0.77	0.80	0.79	0.46	0.51
4	0.70	0.72	0.72	0.37	0.39
5	0.63	0.65	0.66	0.30	0.30
6	0.56	0.57	0.60	0.26	0.22
7	0.50	0.50	0.55	0.22	0.15
8	0.44	0.44	0.50	0.19	0.10

Table 5.2b. Same as Table 5.2a for Blue Canyon

Time Shift	Obs	eWind	SynFcst	eWind Error	SynFcst Error
0	1	1	1	1	1
1	0.93	0.95	0.92	0.77	0.82
2	0.86	0.89	0.85	0.60	0.69
3	0.80	0.83	0.79	0.48	0.57
4	0.74	0.77	0.73	0.39	0.48
5	0.68	0.71	0.68	0.33	0.40
6	0.63	0.66	0.63	0.28	0.33
7	0.58	0.61	0.59	0.24	0.27
8	0.53	0.57	0.55	0.21	0.22

Table 5.2c. Same as table 5.2a for Storm Lake

Time Shift	Obs	eWind	SynFcst	eWind Error	SynFcst Error
0	1	1	1	1	1
1	0.92	0.94	0.91	0.73	0.76
2	0.85	0.88	0.84	0.53	0.59
3	0.78	0.83	0.78	0.41	0.46
4	0.71	0.77	0.72	0.32	0.35
5	0.65	0.72	0.66	0.26	0.27
6	0.60	0.67	0.61	0.21	0.20
7	0.54	0.62	0.56	0.18	0.15
8	0.49	0.57	0.52	0.15	0.12

Table 5.2d Same as Table 5.2a for Lake Benton

Time Shift	Obs	eWind	SynFcst	eWind Error	SynFcst Error
0	1	1	1	1	1
1	0.92	0.95	0.91	0.79	0.82
2	0.84	0.89	0.85	0.62	0.68
3	0.77	0.83	0.79	0.51	0.57
4	0.70	0.77	0.73	0.42	0.47
5	0.65	0.72	0.68	0.36	0.40
6	0.59	0.66	0.63	0.32	0.34
7	0.54	0.61	0.58	0.28	0.28
8	0.49	0.57	0.54	0.24	0.23

Last, we considered the correlation of forecast errors between projects, and particularly as a function of distance between them. The correct modeling of the spatial correlation of forecast errors was an important consideration for this study as it looks at the aggregate impact of many wind projects over a large region. If the synthesized forecast errors are not correlated enough between projects, then the aggregate forecast error will be underestimated, and therefore also the impacts of those errors on system operations; overestimating the degree of correlation between projects will have the opposite effect. The results are presented in the table below. The synthesized forecasts exhibit a greater degree of correlation of forecast errors for a given distance, indicating a possible tendency to overestimate the errors when aggregated over large numbers of projects in a given region. However, the difference is small and the functional dependence with distance is similar.

Table 5.3. Correlation of forecast errors as a function of distance between project pairs

Distance (km)	Output	eWind Error	SynFcst Error
168	0.61	0.21	0.27
303	0.62	0.21	0.27
913	0.23	0.03	0.08
1047	0.15	0.03	0.06
1214	0.16	0.03	0.05
1342	0.11	0.04	0.03

6 One-Minute Output

In the final task, AWS Truewind simulated one-minute plant output data for the onshore and offshore sites. The data were produced for X time windows of X hours each selected by NREL.

To produce the data, AWS Truewind employed a computer program to sample four-hour windows of historical one-minute data from existing wind projects. The source of the samples

was two years of one-minute plant output data provided in a previous project by ERCOT for 17 substations serving seven Texas wind projects.² The program removed ten-minute trends from the data using a bicubic fitting procedure, and then added the residuals to the simulated ten-minute output for each site. It did not allow the same window of residuals to be applied to two different sites in the same time period, as this would have resulted in perfect correlation of the one-minute fluctuations between those sites, whereas in reality one-minute fluctuations between wind projects are entirely uncorrelated. The program excluded from the training data one-minute changes greater than 5% of the plant rated capacity, as they correspond to plant outages, curtailments, and restarts unrelated to the wind.

Figure 6.1 shows a typical sample of one-minute simulated and actual plant output data for a single project overlaid on hourly data for the same site. The synthesized one-minute data display an obvious similarity to the actual data. Both charts show the wide deviations that can occur within an hour due to the passage of weather fronts and turbulent fluctuations in wind speed.

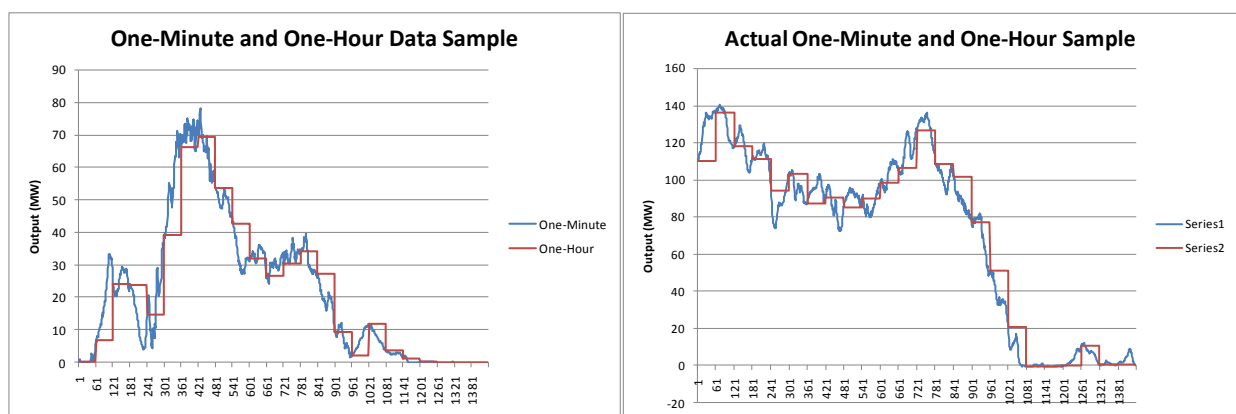


Figure 6.1. Sample of one-minute data overlaid on corresponding one-hour data. The left-hand chart shows simulated data for one of the sites, the right-hand chart shows actual data from a Texas wind project.

Figure 6.2 compares the distribution of step changes of one-minute plant output for an existing project in Texas and for the simulated data for the same location. As with the one-hour step changes, the distributions are superficially very similar. A closer examination shows that the actual plant output rose or dropped in one minute by amounts greater than 5% of the rated capacity of the project more often than a normal distribution of changes would allow.³ In the largest such change at this project, the output increased from zero to 80.5 MW in one minute. Almost certainly this indicates the restoration of the grid connection after an outage or curtailment rather than a surge of production caused by a project-wide wind gust. Analysis of the output data suggests that almost all one-minute step changes greater than 5% of rated capacity are not caused by wind fluctuations but by some other factor, such as strings of turbines going on or off line. Such events were therefore excluded from the training sample in creating the one-minute wind output data.

² The files provided by ERCOT contained data for 32 substations. However, 17 of these did not have valid data. Of the remaining 15 substations, several represented different parts of the same project.

³ At this project, the standard deviation of step changes, as a fraction of rated capacity, is about 0.9%. A 5% change is therefore more than 5 standard deviations. The probability of such an event occurring by chance, assuming a normal distribution of changes, is about 1×10^{-7} , implying one such event in 20 years. Approximately 1300 such changes were actually observed at this project in 2005.

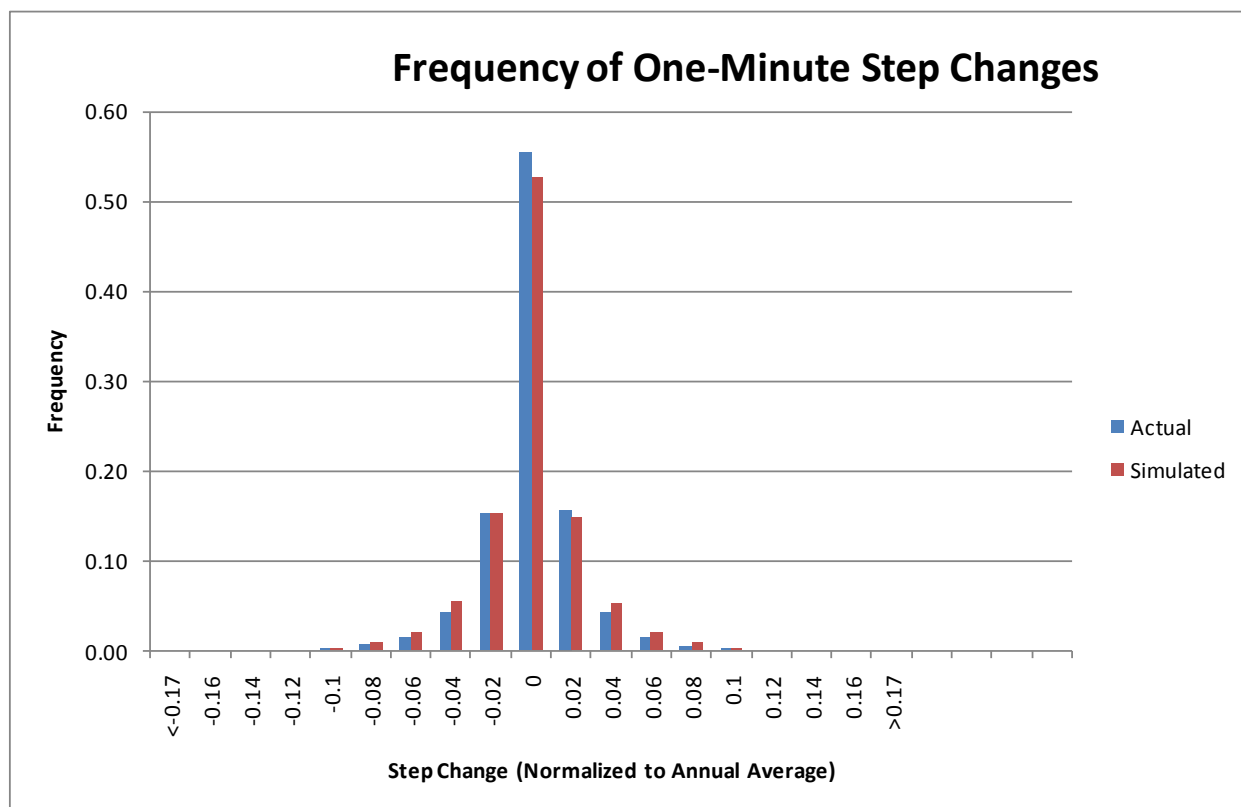


Figure 6.2. Frequency distribution of step changes of simulated and actual one-minute output for a single wind project.

APPENDIX VALIDATION REPORT